Methods used to Verify the 1999 CCFP Forecasts

Statistical results are computed by the Real-Time Verification System (RTVS; Mahoney et al. 1997), developed by the Forecast Systems Laboratory with funds provided by the Federal Aviation Administration Aviation Weather Research Program (Sankey et al. 1997). Statistical results for the 1999 CCFP is described in Mahoney et al. 2000a and Mahoney et al. 2000b.

1. Description of the Forecast Products

Collaborative Convective Forecast Product (CCFP) - This experimental forecast was generated from input provided by participating airline, Center Weather Service Units (CWSU) and Aviation Weather Center (AWC) meteorologists, and staff at the FAA Air Traffic Control System Command Center. The CCFP product was generated as a graphic depicting forecasts of convective activity valid at specific times. Forecasts were issued at 1500 and 1900 UTC with 1-, 3-, and 5-h and 3-, 5-, 7-h lead times, respectively. The forecast product was ultimately used by decision-makers for routing traffic around convective areas (Phaneuf and Nestoros 1999).

First Guess Forecast (FG; also referred to as the Preliminary Forecast) - The FG forecast (Phaneuf and Nestoros 1999) was generated by AWC meteorologists as a precursor to the CCFP. The forecast was made available to the CCFP participants, who evaluated the forecast and provided feedback that was ultimately incorporated into the CCFP. Forecasts were issued at 1400 and 1715 UTC with valid times of 1600, 1800, and 2000 UTC and 2200, 0000, and 0200 UTC, respectively.

2. Verifying Observations

Lighting data, radar reflectivity, and the National Convective Weather Detection Product (NCWD), used to verify the CCFP and FG forecasts are described here.

Lightning data were obtained from the National Lightning Data Network (NLDN; Orville 1991). These data include information regarding the locations (latitude and longitude) and times of specific lightning strikes. The lightning observations were used alone and in combination with radar data to infer areas of active convection for verification of the forecasts.

Radar reflectivity (dBZ) fields, available on a 4-km grid, were used as a second type of observed convective field. A threshold of 40 dBZ was used to define areas of convection.

Finally, the NCWDP (Mueller et al. 1997) combines a 2-dimensional mosaic of radar reflectivity with radar-derived cloud top data and a grid of lightning detections from the NLDN. The cloud top data primarily are used to remove anomalous propagation and ground clutter, and the lightning data help to keep the NCWDP current, since lightning data have a lower latency than radar data. The NCWD fields were available on a 4-km grid, with convective storms delineated by a threshold of 40 dBZ, or more than 3 lightning strikes in 10 minutes.

3. Mechanics

Before forecasts were matched to observations, a 20-km grid was laid over the observation field. Each box on the overlay grid was assigned a *Yes* or *No* value depending on whether a positive observation fell within the 20-km box. For each 20-km box, the criteria used in this study to define a positive observation for each type of verification observation included: 1) 4 strikes of lightning in the 20-km box, 2) one 4-km box of radar reflectivity greater than 40 dBZ that fell in the 20-km box, and 3) one 4-km box of NCWDP with a dBZ greater than 40 that fell in the 20-km box. The same procedure was performed for the forecasts, with a 20-km box labeled with a *Yes* forecast when any part of the forecast polygon intersected that box. If a forecast polygon did not intersect the 20-km box, then a *No* forecast was assigned to that box. Other grid sizes have been tested. Results are presented in Mahoney et al. 2000b.

A filter was applied to the NCWD observations in an attempt to screen out isolated short-lived convection. In this case, a 20-km box was assigned a *Yes* observation only when 12 or more 4-km NCWD boxes meeting the 40 dBZ and greater criteria were activated. Otherwise, a *No* observation was assigned to the 20-km box.

Once this process was complete, each box on the 20-km observation grid was matched to each 20-km box on the forecast grid. This technique produced the forecast/observation pairs used to generate the verification statistics. For example, a *Yes* forecast box and a *Yes* observation box would produce a *Yes*-*Yes* pair. Similarly, a *Yes* forecast and *No* observation would produce a *Yes-No* pair, and so on, filling in the four cells of the statistical contingency table (described further in Section 4).

Observations that fell within a 10-minute time window prior to the forecast valid time were mapped to the 20-km grid and used for verification.

The forecasting domain defined for the evaluation extends west from the Atlantic Ocean to a north-south line east of Denver, Colorado (Fig. 1).

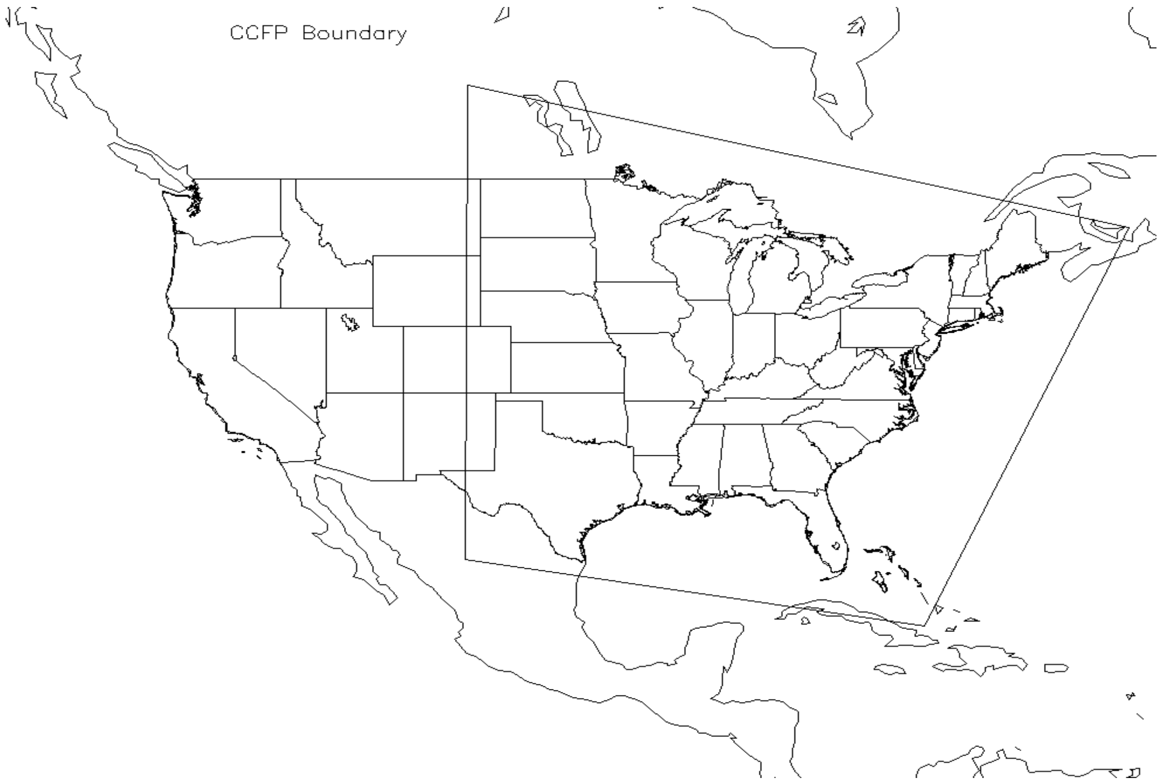

Figure 2. Solid line represents geographic boundary defined for the exercise.

4. Verification Methods and Stratifications

Verification methods are based on standard verification concept (Murphy and Winkler 1987 and Brown et al. 1997). The *Yes/No* forecast/observation pairs were used to create counts, to fill in a 2x2 contingency table like the one shown in Table 1 (Brown et al. 1997). That is, for a given forecast, all of the 20-km boxes with a *Yes* forecast and a *Yes* observation were counted to obtain YY; all of the 20-km boxes with a *Yes* forecast and a *No* observation were counted to obtain YN; and so on. Individual forecast contingency tables were accumulated to obtain tables representing particular days, months, or other periods (including the entire forecast period).

Table 1. Basic contingency table for evaluation of dichotomous (e.g., Yes/No) forecasts. Elements in the cells are the counts of forecast-observation pairs.

	Observation		
Forecast	***Yes***	***No***	***Total***
Yes	YY	YN	YY+YN
No	NY	NN	NY+NN
Total	YY+NY	YN+NN	YY+YN+NY+NN

Table 2 lists the verification statistics that were included in the evaluation, with PODy, PODn, and FAR representing the basic verification statistics. General descriptions of these statistics include the following:

- PODy and PODn are estimates of the proportions of *Yes* and *No* observations, respectively, that were correctly forecast (e.g., Brown et al. 1997).
- FAR is the proportion of *Yes* forecasts that were incorrect.
- Bias is the ratio of the number of *Yes* forecasts to the number of *Yes* observations, and is a measure of over- or under-forecasting.
- The Critical Success Index (CSI), also known as the Threat Score, is the proportion of hits that were either forecast or observed.
- The True Skill Statistic (TSS) (e.g., Doswell et al. 1990) is a measure of the ability of the forecast to discriminate between *Yes* and *No* observations; TSS also is known as the Hanssen-Kuipers discrimination statistic (Wilks 1995).
- The Heidke Skill Score (HSS) is the percent correct, corrected for the number expected to be correct by chance.
- The Gilbert Skill Score (GSS) (Schaefer 1990), also known as the Equitable Threat Score, is the CSI corrected for the number of hits expected by chance.
- The % Area is the percent of the total possible area that had a *Yes* forecast (Brown et al. 1997).

Table 2. Verification statistics used in this study.

Statistic	***Definition***	***Description***
PODy	YY/(YY+NY)	Probability of Detection of “Yes” observations
PODn	NN/(YN+NN)	Probability of Detection of “No” observations
FAR	YN/(YY+YN)	False Alarm Ratio
CSI	YY/(YY+NY+YN)	Critical Success Index
Bias	(YY+YN)/(YY+NY)	Forecast Bias
TSS	PODy + PODn – 1	True Skill Statistic
HSS	[(YY+NN)-C1]/(N-C1), where N=YY+YN+NY+NN C1=[(YY+YN)(YY+NY) + (NY+NN)(YN+NN)] / N	Heidke Skill Score
GSS	(YY-C2)/[(YY-C2)+YN+NY], where C2=(YY+YN)(YY+NY)/N	Gilbert Skill Score
% Area	(Forecast Area) / (Total Area) x 100	% of the area of the continental U.S. where convection is forecast to occur

The convective areas defined by the CCFP were stratified using 4 types of criteria: maximum tops (e.g. height), areal coverage, probability of occurrence, and growth rate. The statistical results were also stratified using these categories. The stratification criteria and their categories are:

1) Maximum Tops (Height)
- At or above 25,000 ft;
- 25 - 31,000 ft;
- 31 - 37,000 ft; and
- Above 37,000 ft.

2) Areal Coverage. Statistics were generated for each of the coverage categories, however, no attempt was made to vary the observation criteria within a specific coverage category.

From 1 - 26 June; the coverage categories were
- 25% and above;
- 25 - 49%; and
- 50% and above.

From 27 June - 31 August, the coverage categories
- 25% and above;
- 25 - 49%;
- 50 - 74%; and
- 75% and above.

3) Probability of Occurrence
- High (70 - 100%);
- Medium (40 - 69%); and
- Low (1 - 39%).

The method for identifying the high, medium, and low probability areas changed from single circles during the evaluation period to circles within circles. As a consequence, statistics were generated independently for each specific probability category. Results are presented on the web-based interface for all categories of height and coverage.

5. References

Brown, B.G., G. Thompson, R.T. Bruintjes, R. Bullock, and T. Kane, 1997: Intercomparison of in-flight icing algorithms. Part II: Statistical verification results. *Wea. and Forec.*, **12,** 890-914.

Doswell, C.A., R.Davies Jones, and David L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. and Forec.*, **5**, 576-585.

Mahoney, J.L., B.G. Brown, and J. Hart, 2000a: Statistical Verification Results for the Collaborative Convective Forecast Product. NOAA Technical Report OAR 457-FSL 6, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Forecast Systems Laboratory, 30 pp.

Mahoney, J.L., B.G. Brown, C. Mueller, and J.E. Hart, 2000b: Convective Intercomparison Exercise: Baseline Statistical Results. Preprints. 9th Conference on Aviation, Range, and Aerospace Meteorology, 11-15 September, Orlando, FL.

Mahoney, J.L., J.K. Henderson, and P.A. Miller, 1997: A Description of the Forecast Systems Laboratory's Real-Time Verification System (RTVS). Preprints, *7th Conference on Aviation, Range, and Aerospace Meteorology*, Long Beach, American Meteorological Society, J26-J31.

Mueller, C.K., C.B. Fidalego, D.W. McCann, D. Meganhart, N. Rehak, and T. Carty, 1999: National Convective Weather Forecast Product. *Preprints, 8th Conference on Aviation Range, and Aerospace Meteorology,* American Meteorological Society (Boston), 230-234.

Murphy, A.H. and R.L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330-1338.

Orville, R.E., 1991: Lightning ground flash density in the contiguous United States-1989. *Mon. Wea. Rev*., **119**, 573-577.

Phaneuf, M. W. and D. Nestoros, 1999: Collaborative convective forecast product: Evaluation for 1999. (Available from the author at CygnaCom Solution, Inc.)

Sankey, D., K.M. Leonard, W. Fellner, D.J., Pace, K.L. Van Sickle, 1997: Strategy and Direction of the Federal Aviation Administration's Aviation Weather Research Program. Preprints, *7th Conference on Aviation, Range, and Aerospace Meteorology*, Long Beach, American Meteorological Society, 7-10.

Schaefer, J.T., 1990: The Critical Success Index as an indicator of warning skill. *Wea. and Forec*., **5**, 570-575.

Wilks, D.S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.